HPC
High-End into the Mainstream

Asia HPC Conference
November 2005

Geoff Lowney

Intel Fellow

Microprocessor Architecture and Planning

China Fellow-in-Residence

intel

Where it all Began

- Eniac I (1946)
  - About 5000 additions per second
  - Cost about $500,000
  - Weighed more than 30 tons
  - Could store 20 numbers in main memory

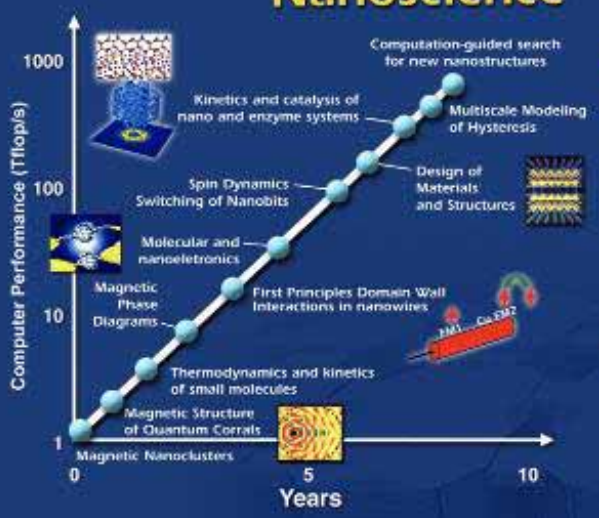
Looking at HPC Today

The Next Milestone

Source: www.top500.org, November 2004

Petaflop Usage Models

Nanoscience



Expected Outcomes

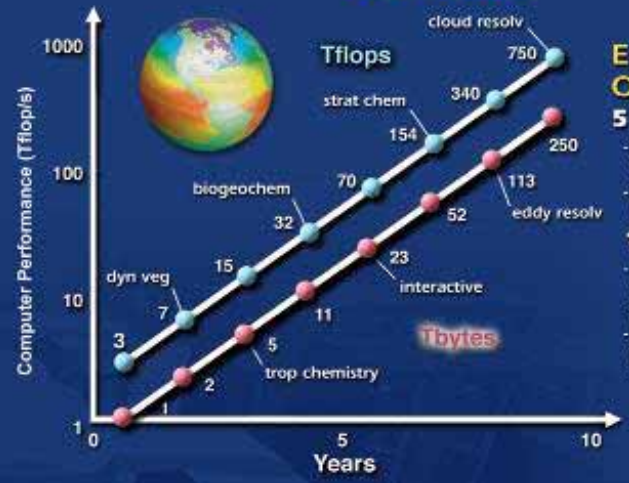
5 years

- Realistic simulation of self-assembly and single-molecule electron transport
- Finite temperature properties of nanoparticles/quantum corrals

10 years

- Multi-scale modeling of molecular electronic devices
- Computation-guided search for new materials/nanostructures

Climate



Expected Outcomes

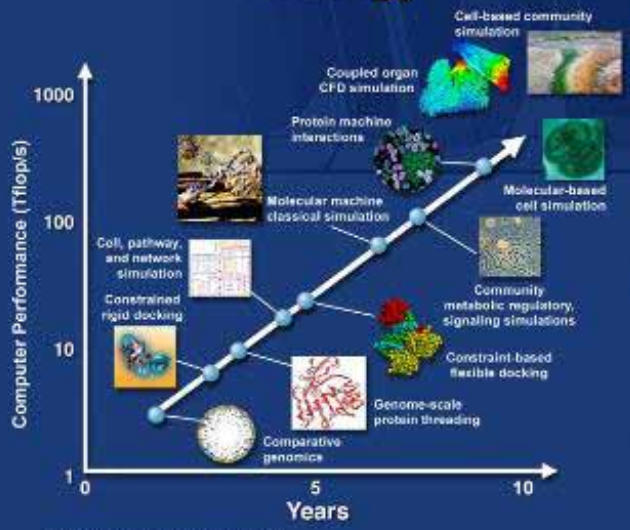
5 years

- Fully coupled carbon-climate simulation
- Fully coupled sulfur-atmospheric chemistry simulation

10 years

- Cloud-resolving 30-km spatial resolution atmosphere climate simulation
- Fully coupled, physics, chemistry, biology Earth system model

Biology



Expected Outcomes

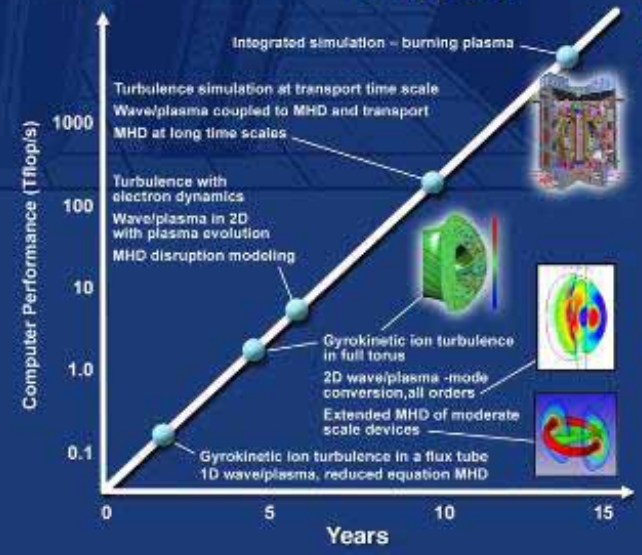
5 years

- Metabolic flux modeling for Hydrogen and Carbon fixation pathways
- Constrained flexible docking simulations of interacting proteins

10 years

- Multi-scale stochastic simulations of combined microbial metabolic, regulatory and protein interaction networks
- Dynamics simulations of complex molecular machines

Fusion



Expected Outcomes

5 years

- Full-torus, electromagnetic simulation of turbulent transport with kinetic electrons for simulation times approaching transport time-scale
- Develop understanding of internal reconnection events in extended MHD, with assessment of RF heating and current drive techniques for mitigation

10 years

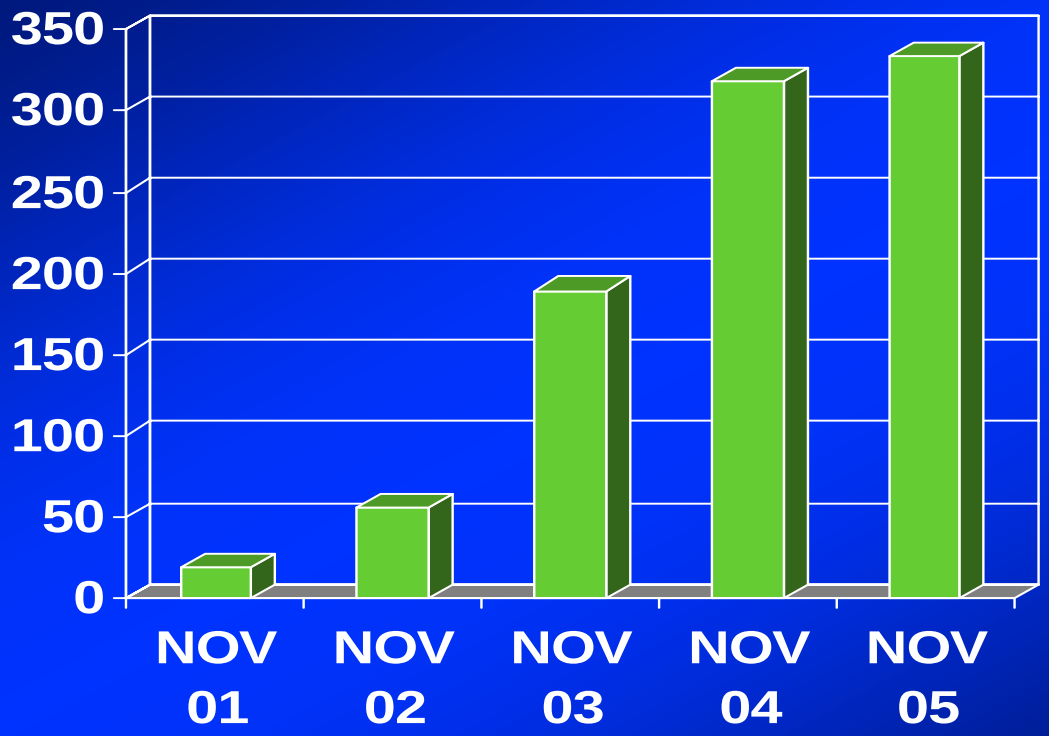
- Develop quantitative, predictive understanding of disruption events in large tokamaks
- Begin integrated simulation of burning plasma devices - multi-physics predictions for ITER

Center for Computational Sciences
Oak Ridge National Laboratory

Office of Science

Intel's Focus on HPC

Intel leads with 333 total systems in TOP500



GEOGRAPHIES

USA – 305

Europe - 100

Asia – 66

Remainder - 29

Source: www.top500.org

intel

Why Do We Focus On HPC?

1 ***Research/Academia HPC*** → 2 ***Enterprise HPC*** → 3 ***Enterprise Business***

Driven by insatiable appetite for compute power

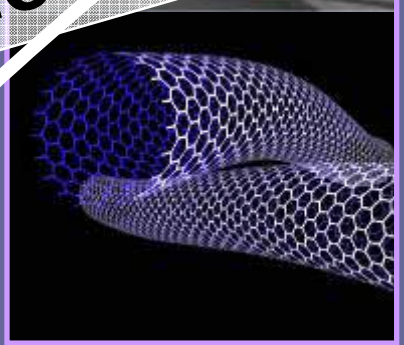
Russia, TeraGrid, EU DataGrid/CERN, China …

Large-scale across multi-geo sites

Mostly open source and in-house apps and middleware

Competitive differentiation

Simulation, data mining and modeling



Large-scale across multi-geo sites

In-house and commercial apps and middleware

Services-Oriented Enterprise

Cross-enterprise functions

Large-scale across the Enterprise

Broad standards-based ecosystem

Technology Diffusion

intel

The Data Explosion

Byte = 8 bits
Kilobyte = 10^3
Megabyte = 10^6
Gigabyte = 10^9
Terabyte = 10^{12}
Petabyte = 10^{15}
Exabyte = 10^{18}
Zettabyte = 10^{21}
Yottabyte = 10^{24}

U.S. Broadcast Media	14,893 TB
Worldwide Filmed Content	420,254 TB
Worldwide Printed Content	1,633 TB

Internet	532,897 TB
World Telephone Calls	17,300,000 TB

Worldwide Magnetic Content	4,999,230 TB
Worldwide Optical Content	103 TB
Electronic Flows Of New Info	17,905,340 TB

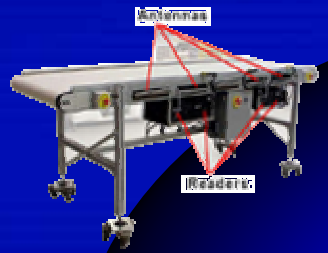
2002 = 5 Exabytes of NEW data
5,000,000,000,000,000,000

intel

Source: How much information 2003
UC Berkeley

The “Mundane” Becomes HPC

WAREHOUSE, SHIP & TRACK



RFID CONVEYOR

RFID

CONTINUAL TRACKING

SMART WAREHOUSE

GPS

SMART VEHICLE LOADING & ROUTING

SMART INVENTORY

intel

Digital Health Becomes HPC

COMPUTATIONAL BIOLOGY

Whole Body Response
Cellular Response
Identify Drug Targets
Protein Structure
Protein Functions in Pathways
Pathways Normal & Aberrant
Protein-Protein Interactions
Map Genes to Proteins
Annotate the Genes
Find the Genes

➢1000
1000
100
10
TeraOps

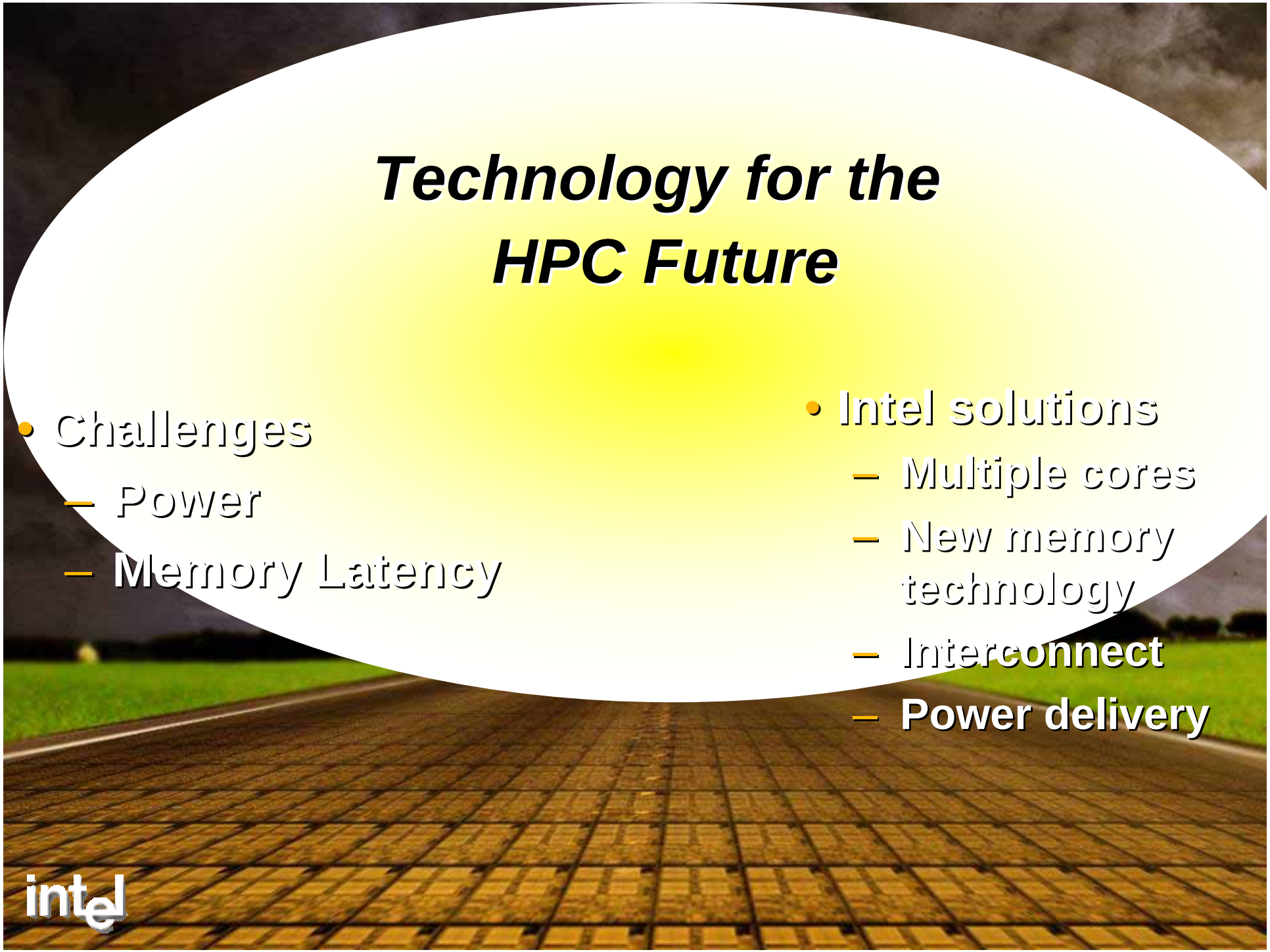
SOURCE: Sandia National Laboratories

REAL-TIME ANALYSIS

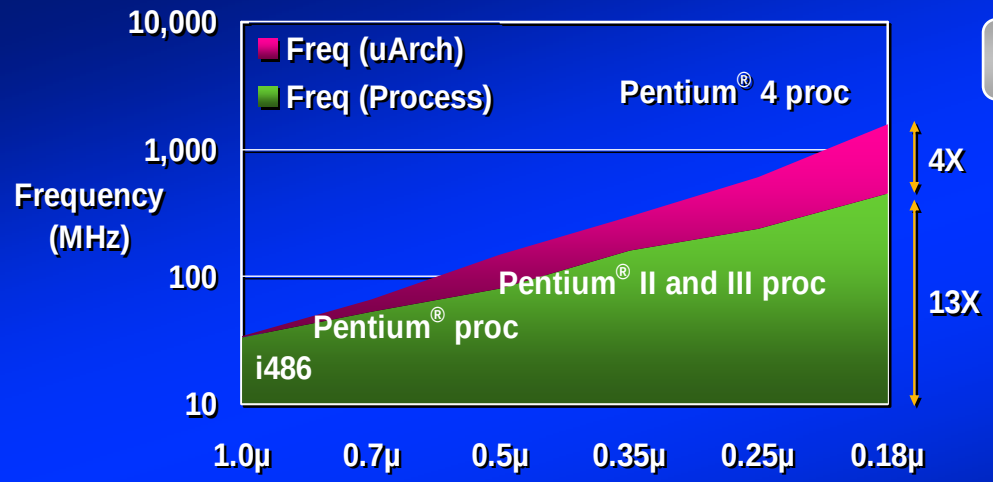
ROBOTIC SURGERY

ADVANCED MEDICINE

intel

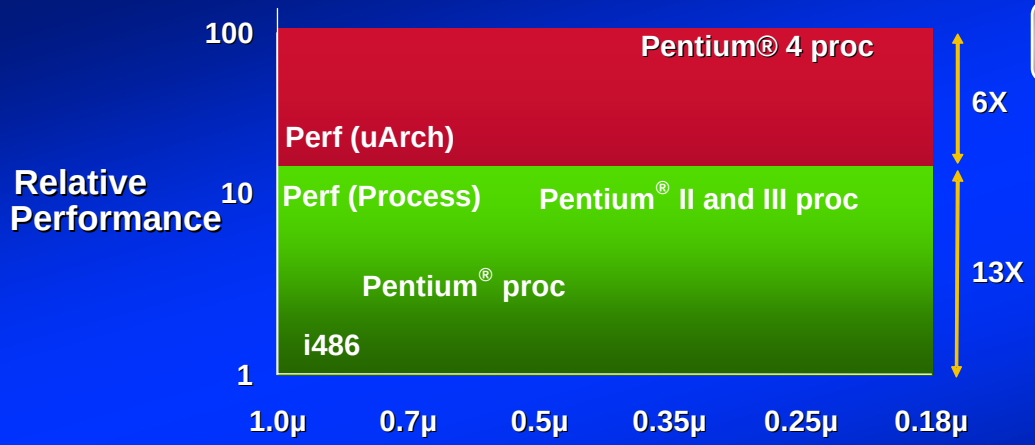

Technology for the HPC Future
Challenges
Power
Memory Latency
Intel solutions
Multiple cores
New memory technology
Interconnect
Power delivery
intel

Historical Frequency & Performance



Frequency Increased 50X

- 13X due to process technology
- Additional 4X due to microarchitecture



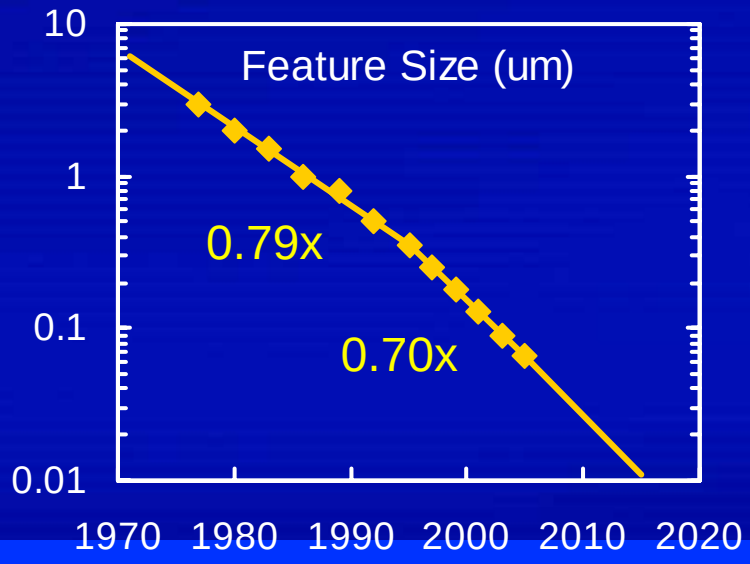
Performance Increased >75X

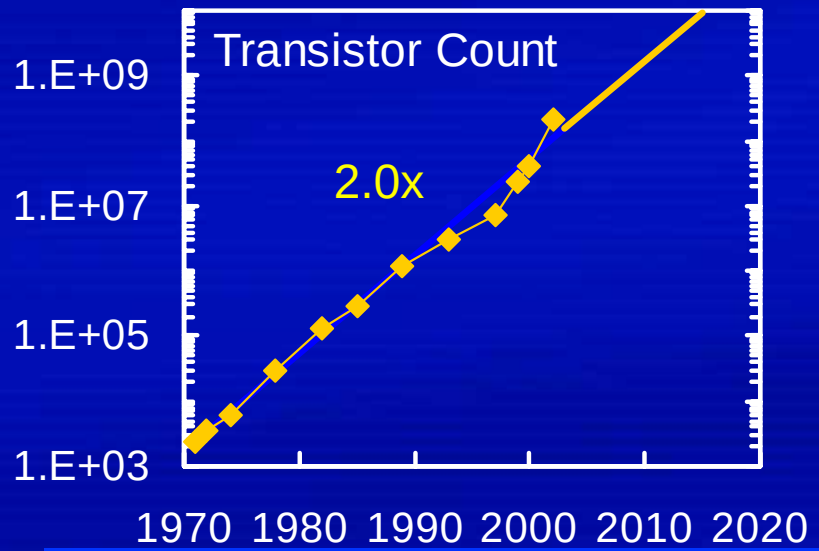
- 13X due to process technology
- Additional >6X due to microarchitecture

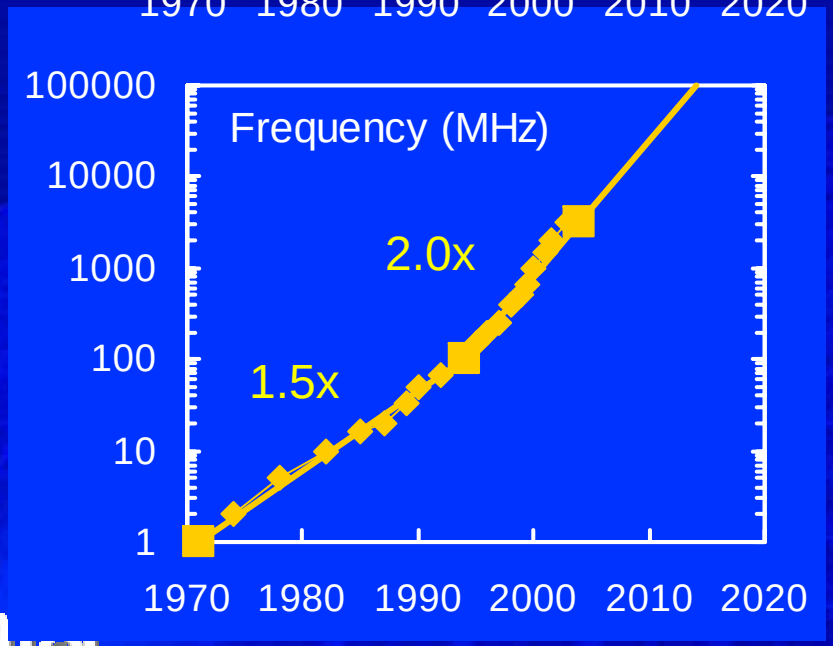
Performance measured using SpecINT and SpecFP

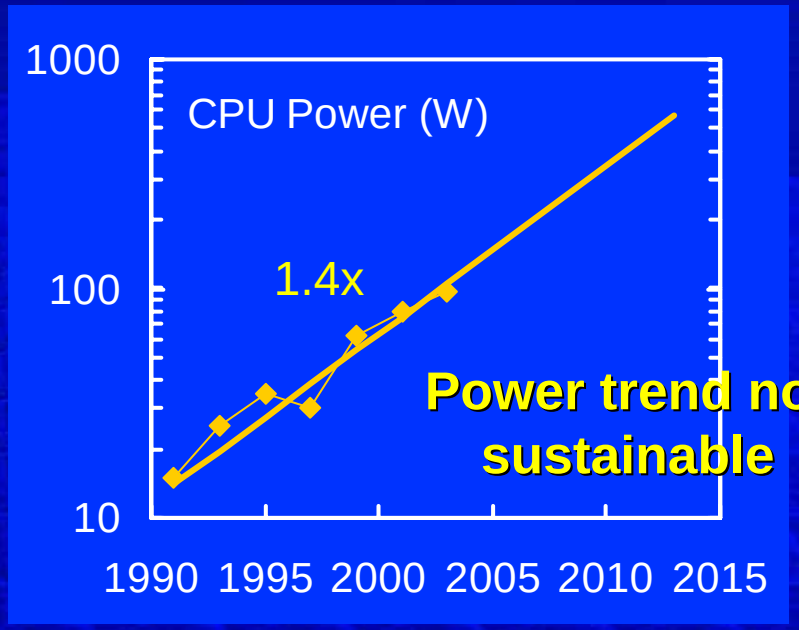
Source: Intel Corporation

Moore's Law at Intel 1970-2005









Reducing power with voltage scaling

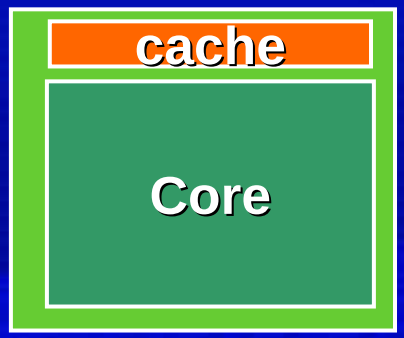
- Power = Capacitance * Voltage**2 * Frequency
- Frequency ~ Voltage in region of interest
- Power ~ Voltage ** 3
- 10% reduction of voltage yields
  - 10% reduction in frequency
  - 30% reduction in power
  - Less than 10% reduction in performance

Rule of Thumb

Voltage	Frequency	Power	Performance
1%	1%	3%	0.66%

Dual Core example of Voltage Scaling

Voltage	Frequency	Power	Performance
1%	1%	3%	0.66%



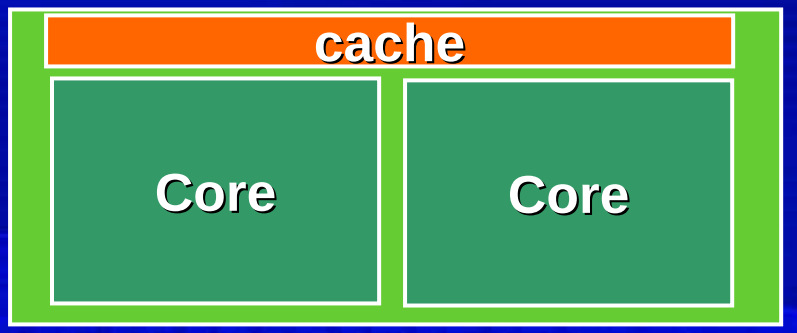
Voltage = 1

Freq = 1

Area = 1

Power = 1

Perf = 1



Voltage = - 15%

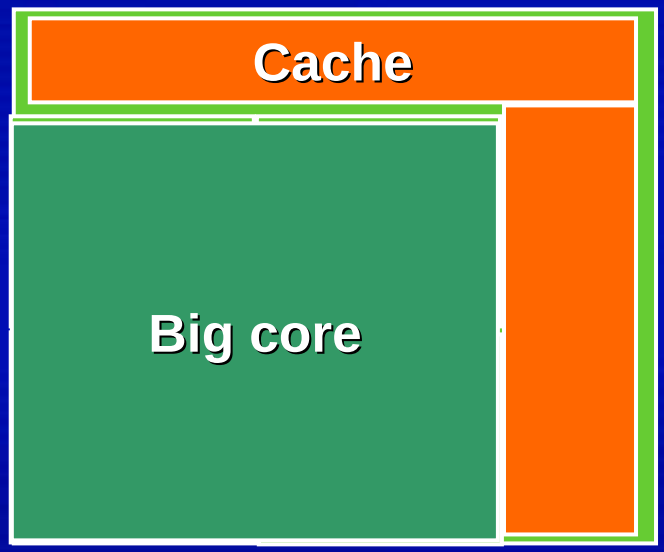
Freq = - 15%

Area = 2

Power = 1

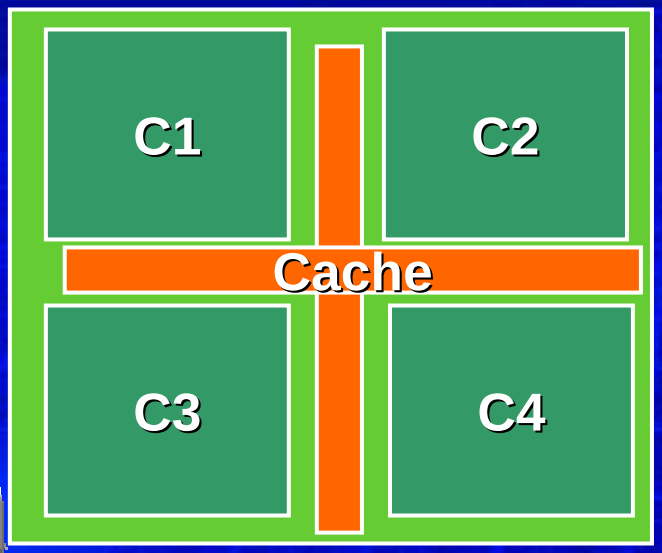
Perf = ~1.8

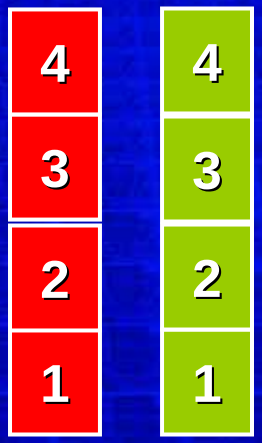
Multiple cores deliver more performance per watt



Power = ¼

Performance = 1/2





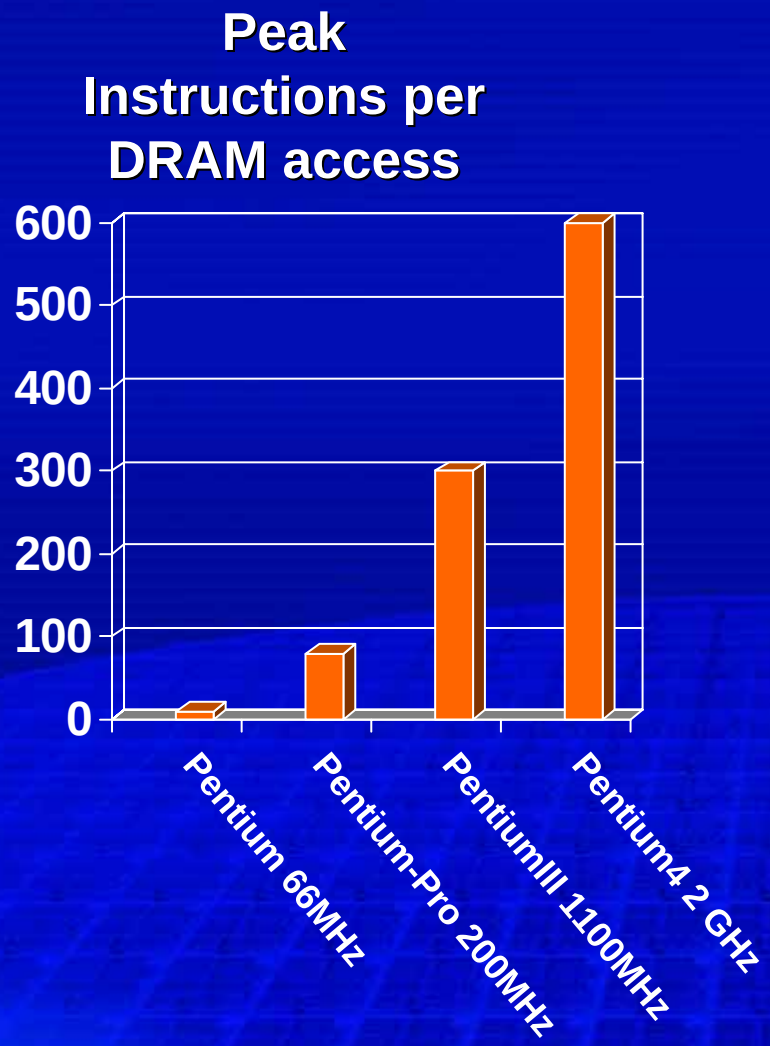
Many core is more power efficient

Power ~ area

Single thread performance ~ area.5**

intel

Memory Latency



- Reduce DRAM access with large caches
  - Extra benefit: power savings. Cache is lower power than logic
- Tolerate memory latency with multiple threads
  - Multiple cores
  - Hyper-threading

intel

Moore's Law will provide transistors

Intel process technology capabilities

High Volume Manufacturing	2004	2006	2008	2010	2012	2014	2016	2018
Feature Size	90nm	65nm	45nm	32nm	22nm	16nm	11nm	8nm
Integration Capacity (Billions of Transistors)	2	4	8	16	32	64	128	256

Use transistors for

- **Multiple cores**
- **On-core memory (caches)**
- **New features (*Ts)**

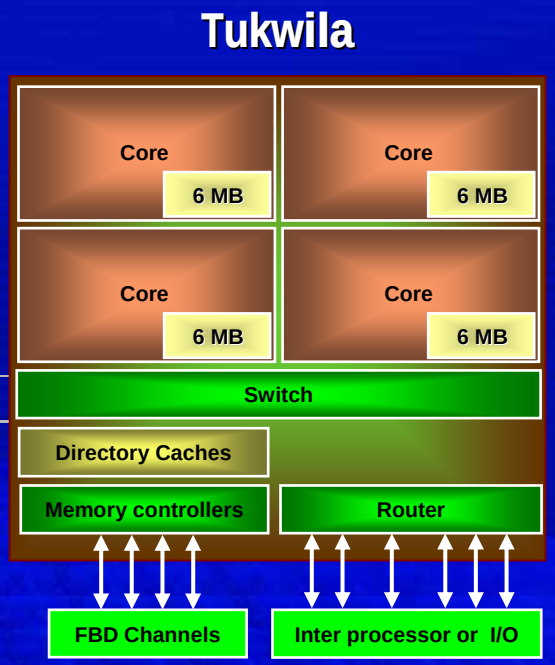
Multiple cores and caches address power and memory latency issues

Taking Us to Teraflop-in-a-Socket

Source: www.top500.org, November 2004

Next Gen Itanium® Microprocessor

- Next Generation Interconnect
- 4 Highly Efficient IPF cores
- Estimated 40G Flops / Socket
- 24M L2 cache
- Server *Ts



NG Interconnect

- Point to point, serial, differential
- Low latency
- 4.8 & 6.4 GT/s/dir, full width
- 4 full width, 2 half width
- Integrated memory controller

Server Technology

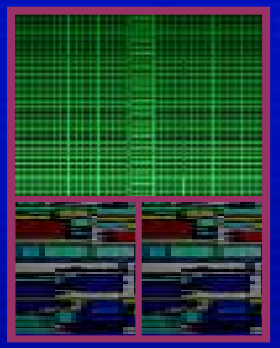
- Virtualization
- Partitioning
- Enhanced RAS
- ~1.3x scalar boost over Montecito

intel

Next Gen Xeon® Microprocessor

- 2 highly efficient OOO cores
- 4M L2 cache
- 40% reduction in TDP power
- Server *Ts

Woodcrest



2H '06

Wider

- Higher FE bandwidth
- 4 wide decode
- 4 wide renaming
- 4 wide retire
- Additional Integer port
- 128 bit wide SSE2 implementation

Deeper Buffers

- Larger RS, ROB
- Larger SB
- More Line Fill buffers (MEM)

More Efficient Pipeline

- Macro-Fusion: CMP+JMP in 1 clock
- Enhanced Micro-op-fusion
- Cache to cache transfer in CMP
- FIFO scheduling in RS
- Pseudo single cyc Branch Predict
- Faster string instructions (REP mov)

intel

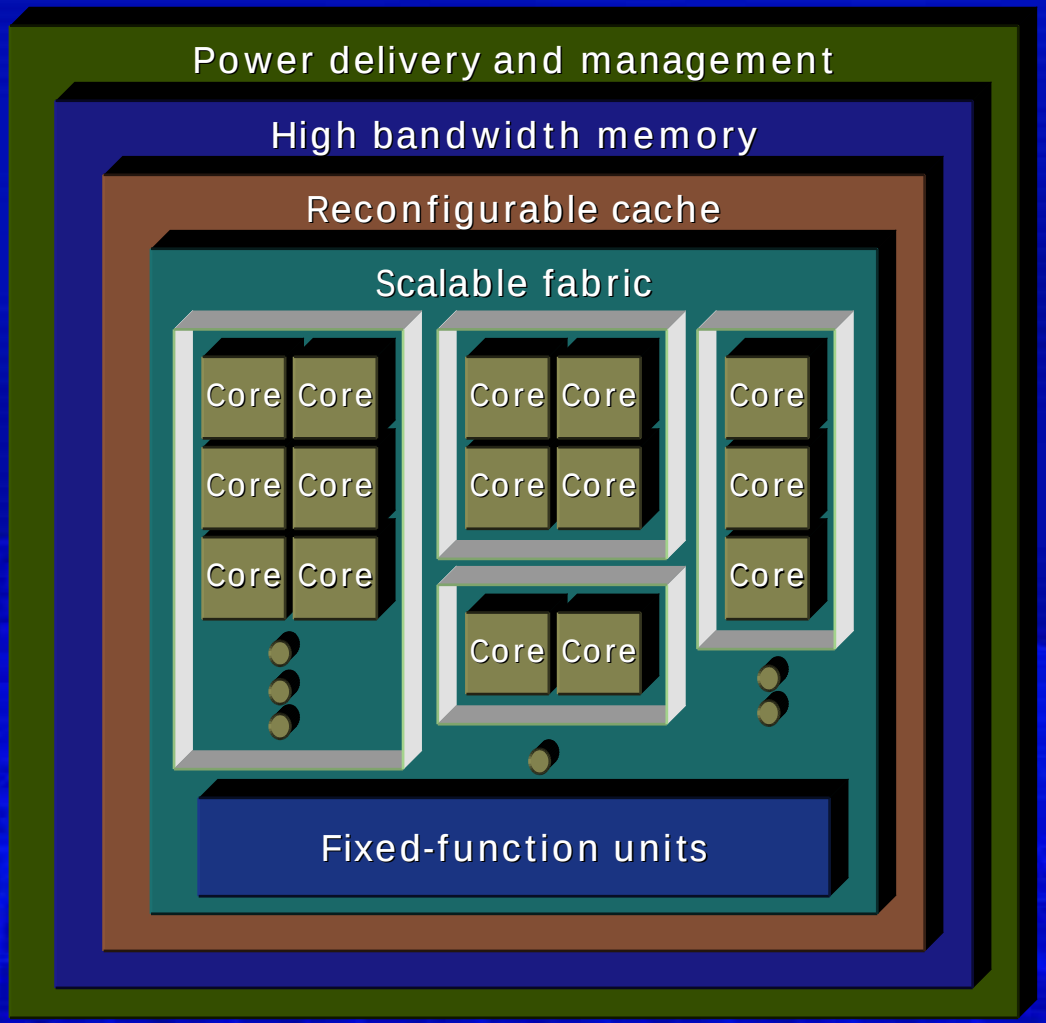
Future Architecture
Many More Cores

Parallel extension of IA

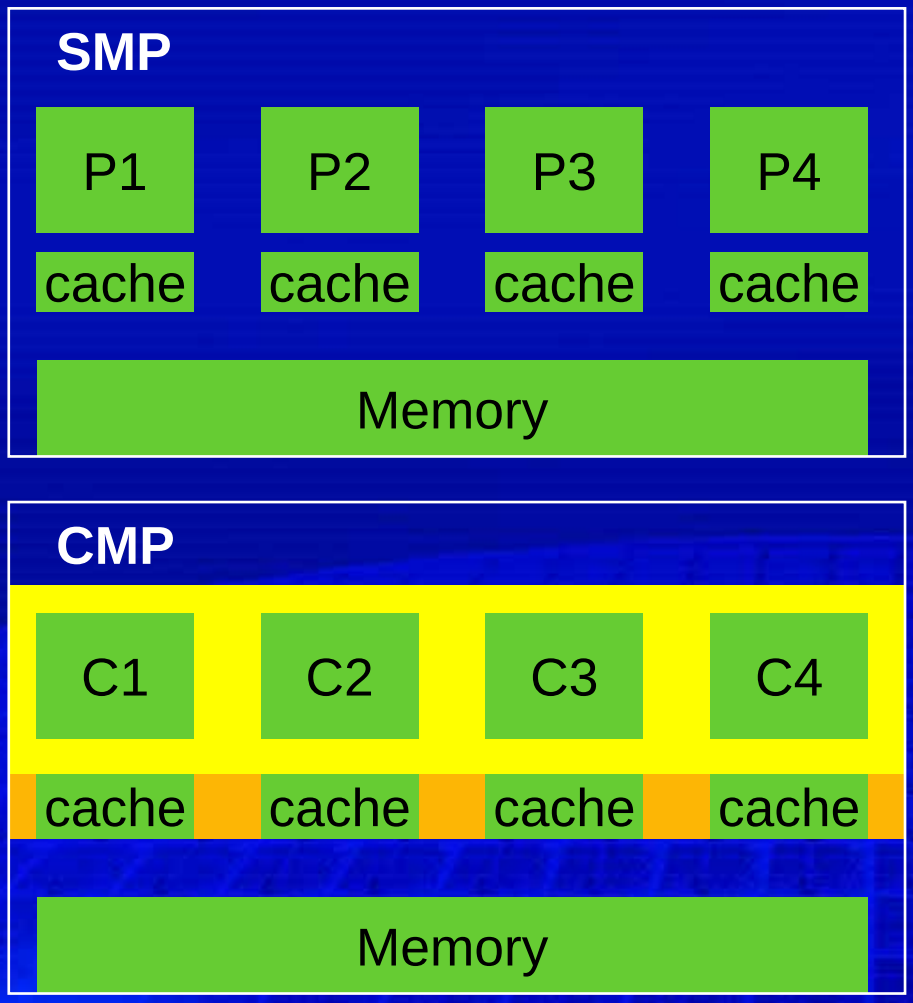
- Homogeneous array of cores
- Fixed-function units
- Coarse- and fine-grained data- and thread-level parallelism
- Global coherency hardware

Partitioned array

- Application domains
- Isolated communication traffic
- Fault tolerance



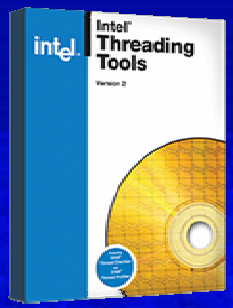
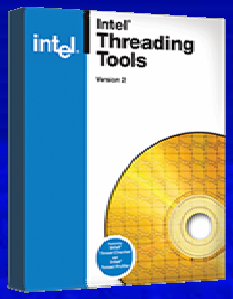
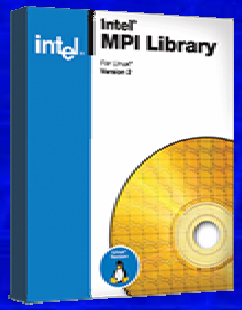
Multiple cores and Parallel Programming



- No change in fundamental programming model
- Synchronization and communication costs greatly reduced
  - Optimization choices may be different
  - Makes it practical to parallelize more programs

intel

Intel Tools for Parallel Programming

	Design / Compile	Debugging	Performance Analysis	Algorithms
Single-Core		Intel® Debugger, Array Visualizer		
Multi-Core	OpenMP* Autoparallel			IPP, MKL already threaded
Cluster				

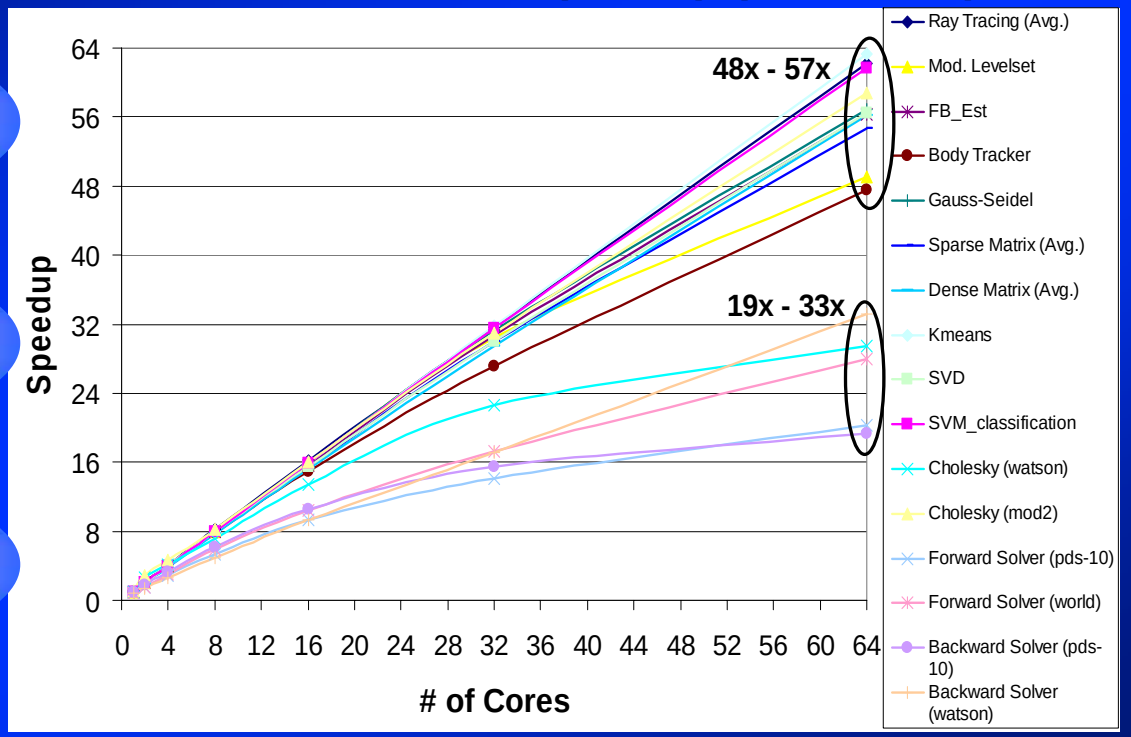
intel

Workload Acceleration

Recognition

Mining

Synthesis



Group I – Scale well with increasing core count

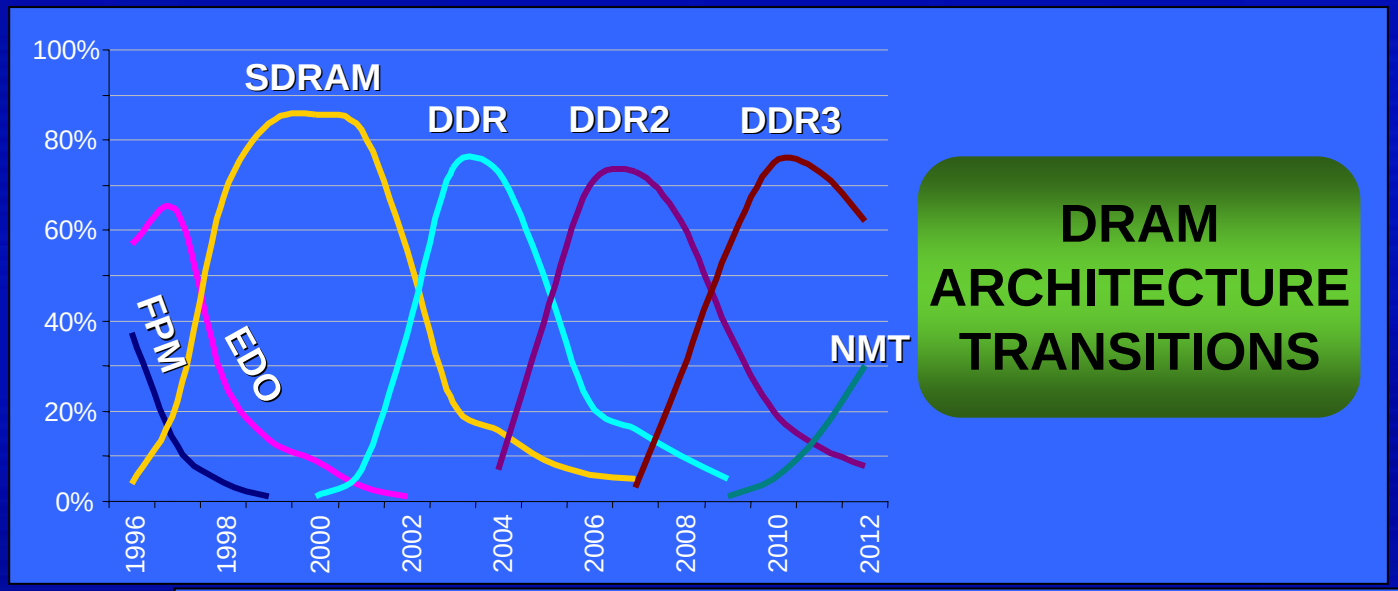
Examples: Ray tracing, body tracking, physical simulation

Group II – Worst-case scaling examples, yet still scale

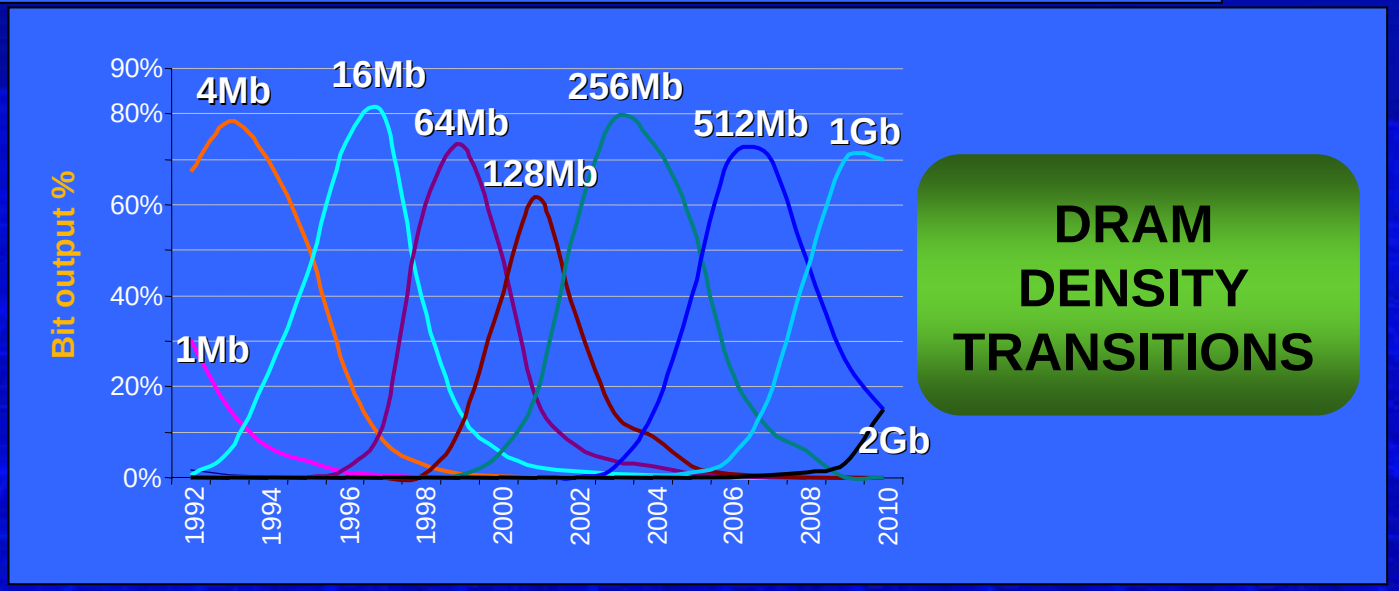
Examples: Forward & backward solvers

Source: Intel Corporation

Feeding the System - Memory



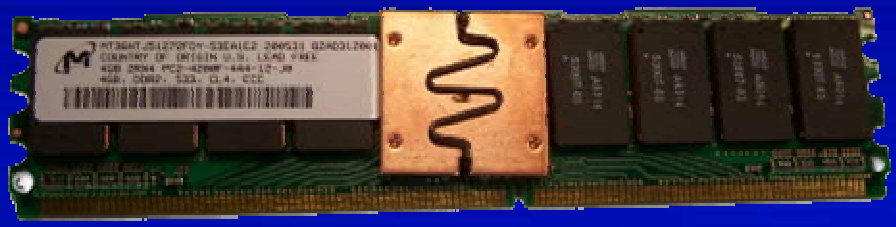
DRAM ARCHITECTURE TRANSITIONS

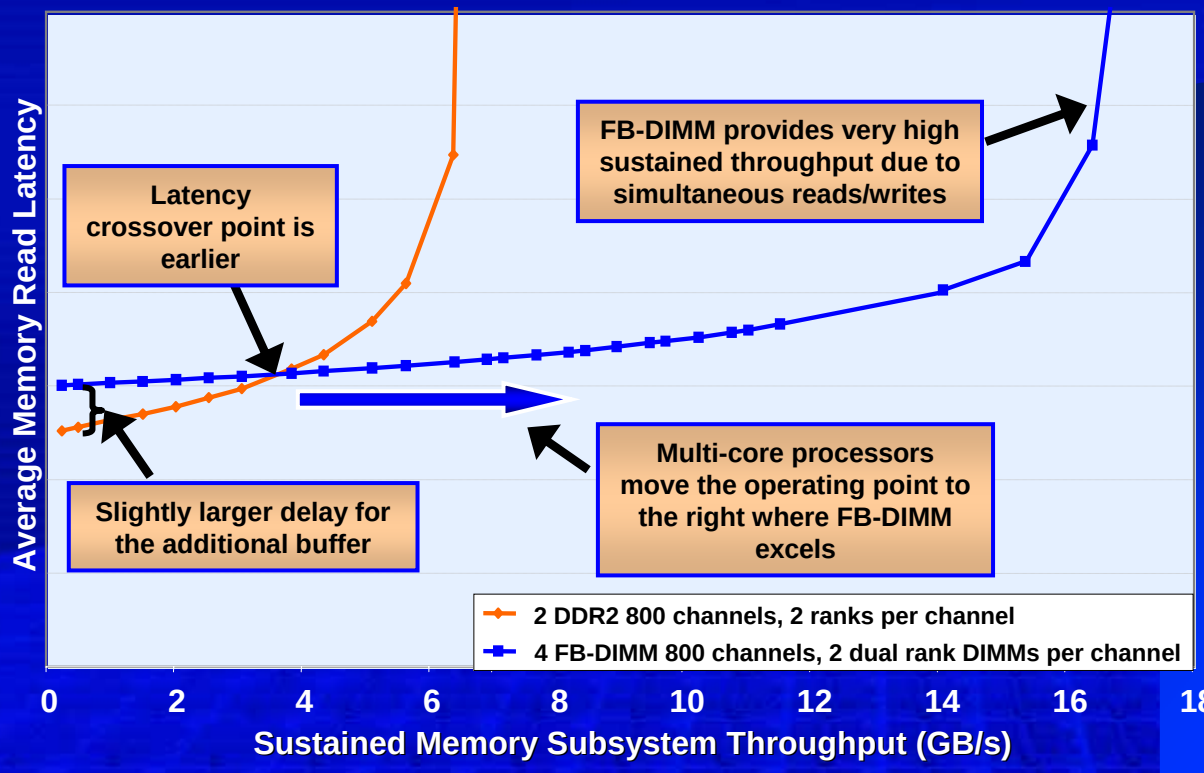


DRAM DENSITY TRANSITIONS

intel

FB-DIMM





Increased bandwidth & capacity

- >3x increase in bandwidth
- 2x the number of channels
- 4x memory capacity

Enhanced Data Availability

- Extended ECC protection
- Memory mirroring - RAID 1
- DIMM sparing & scrubbing

Serial Interface

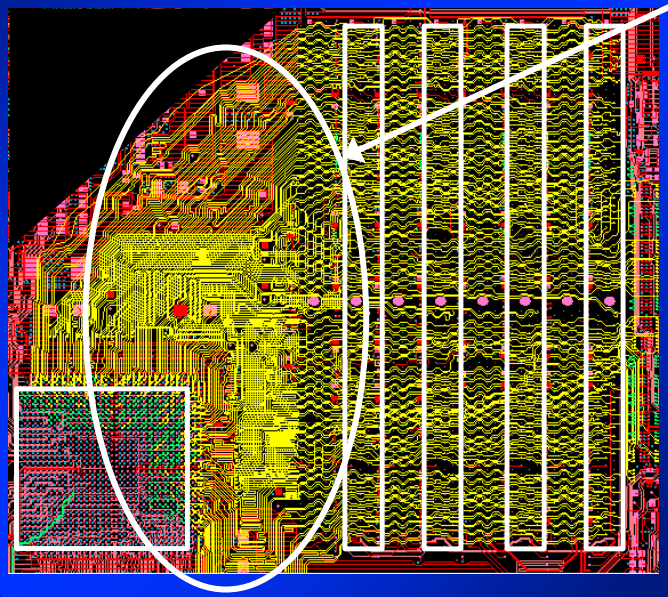
- Lower pin count
- Simplified board routing

Source: Intel Corporation

FBD System Benefits

4 R-DIMMs vs 4 FBDIMMS ROUTING COMPARISON

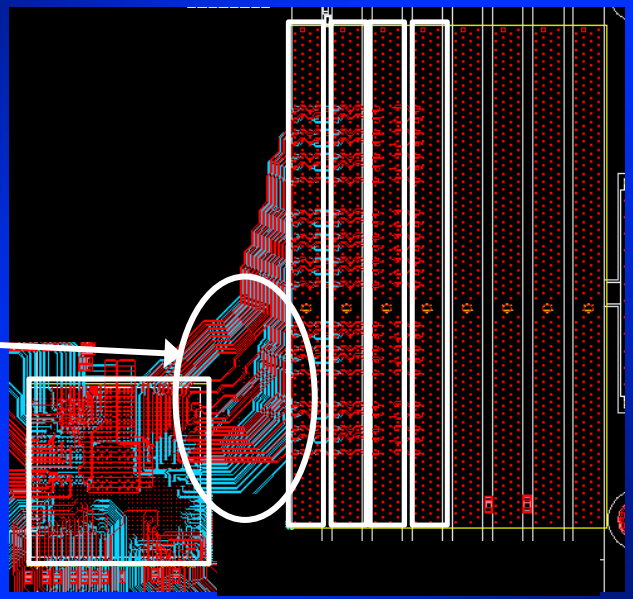
DDR2 R-DIMMs
1 Channel, 2 Routing Layer
3rd layer required for power



Serpentine routing is complicated and uses up a lot of board area

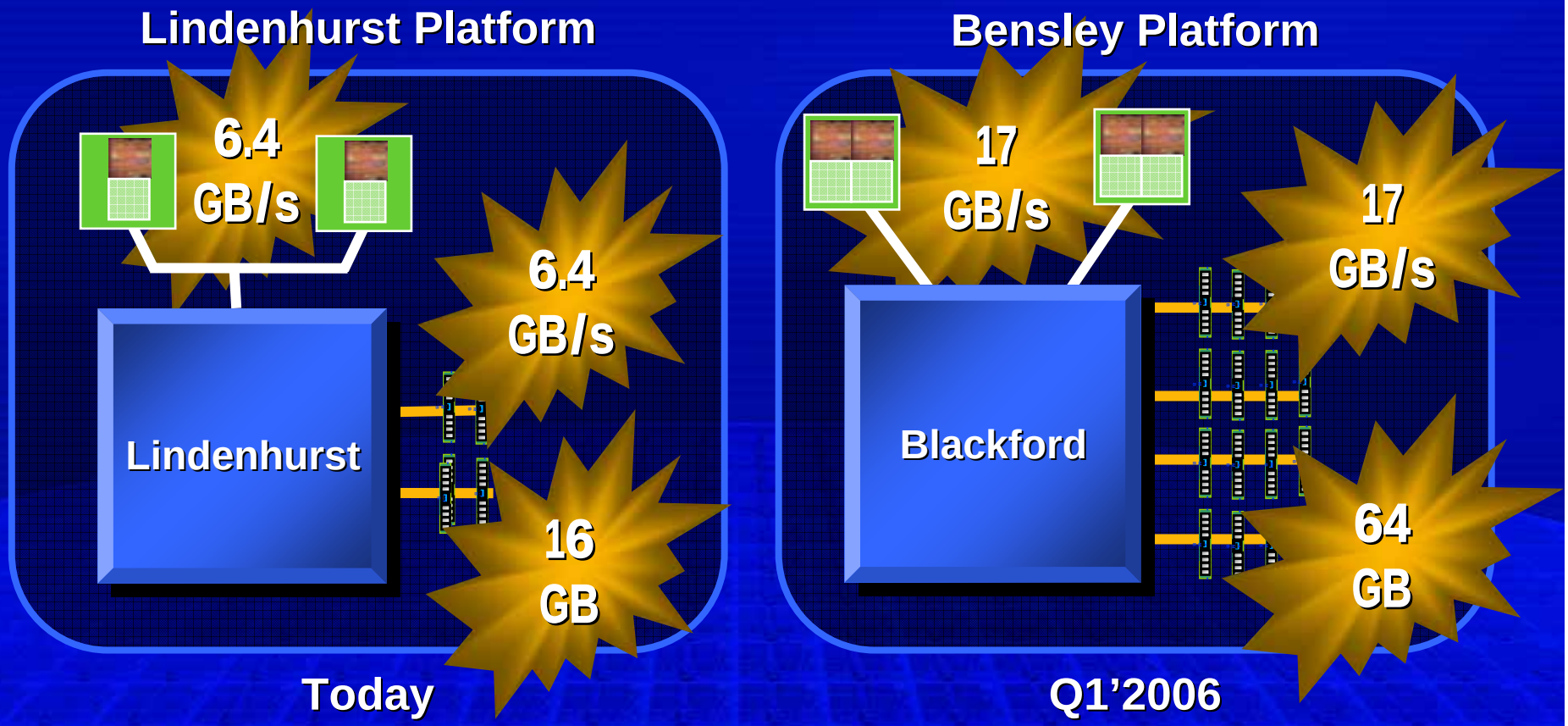
Fewer signals and relaxed trace length matching minimizes board area

FBDs
2 Channels, 2 Routing Layers
Includes power delivery



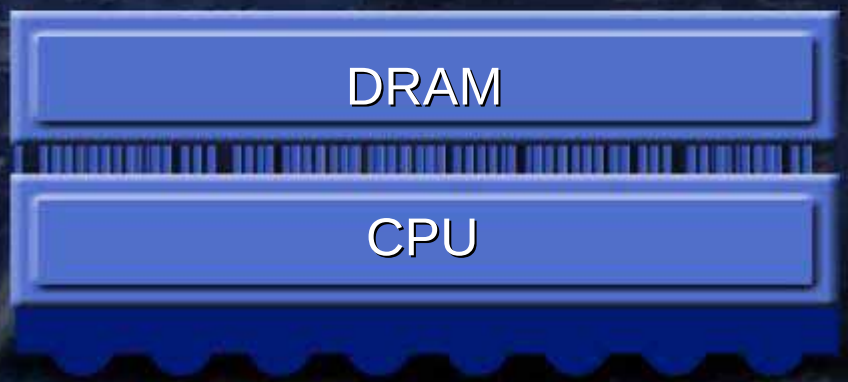
intel

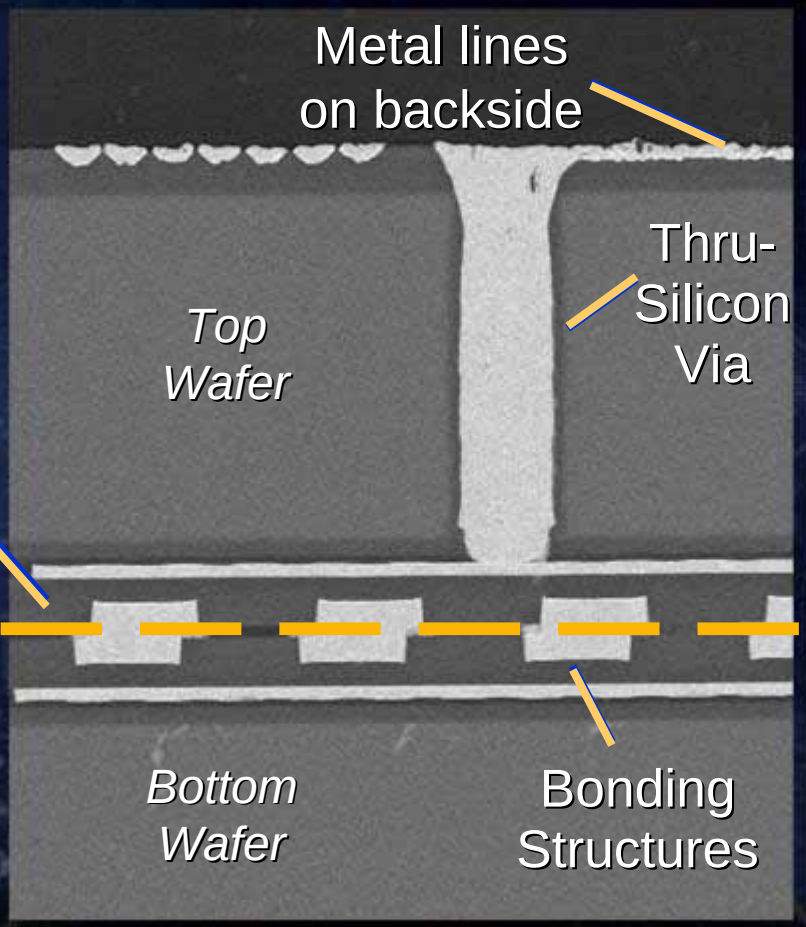
FBD in Next Generation DP Platform



Future Memory Interconnect

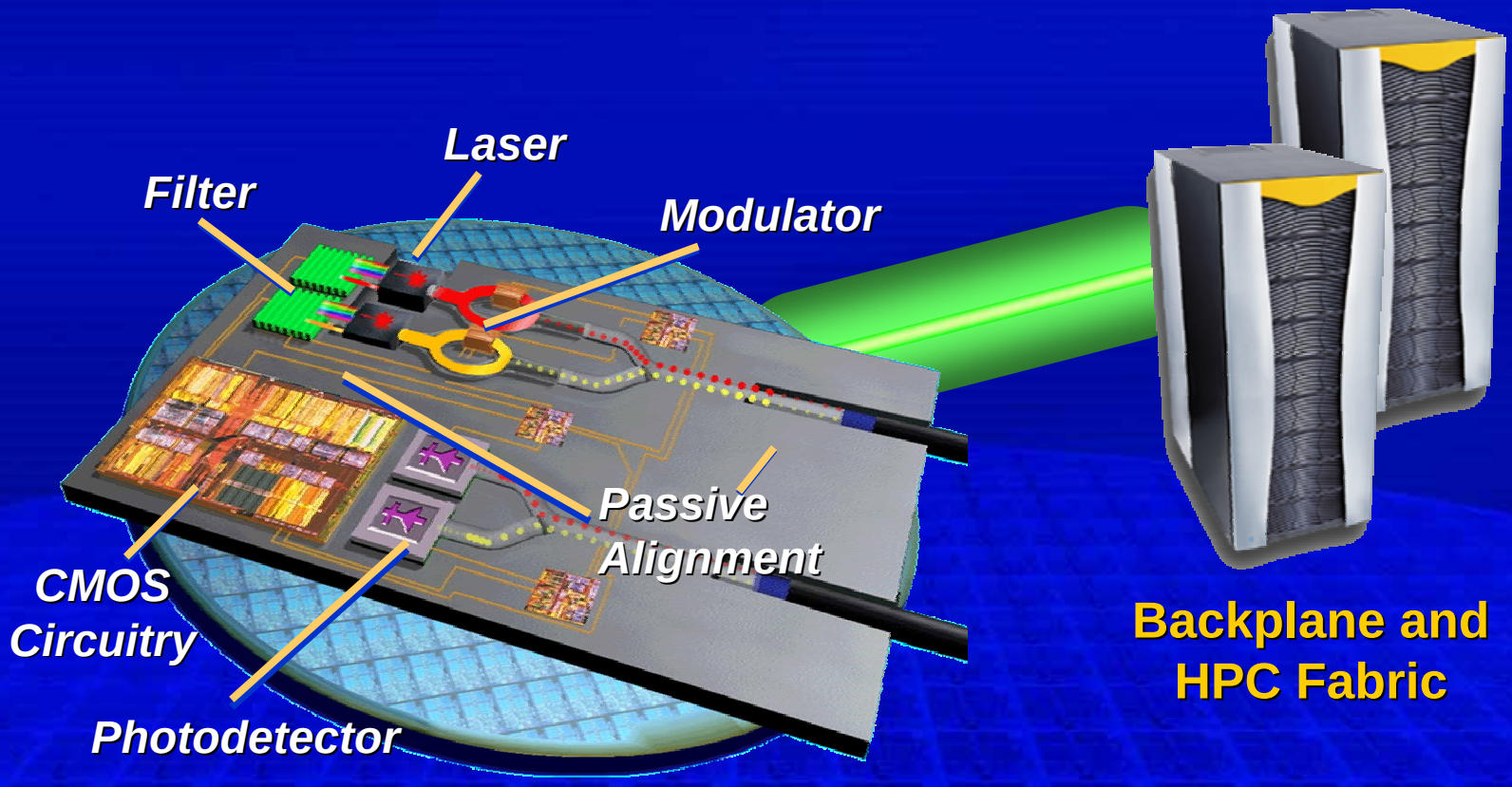
Die Stacking





Source: Intel Corporation

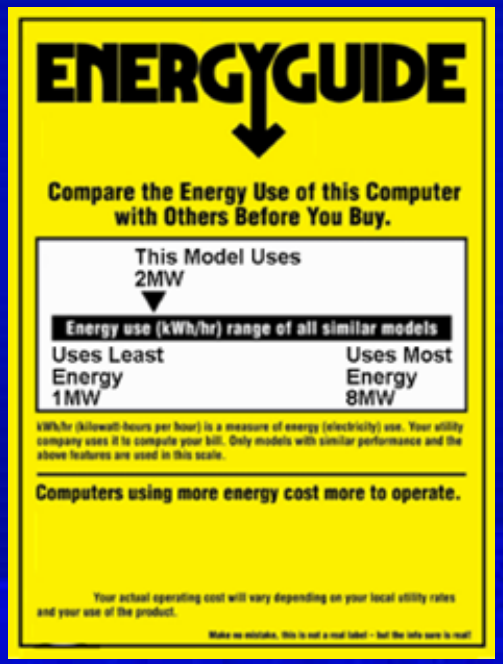
Future System Interconnect
Silicon Photonics



intel

Power Consumption

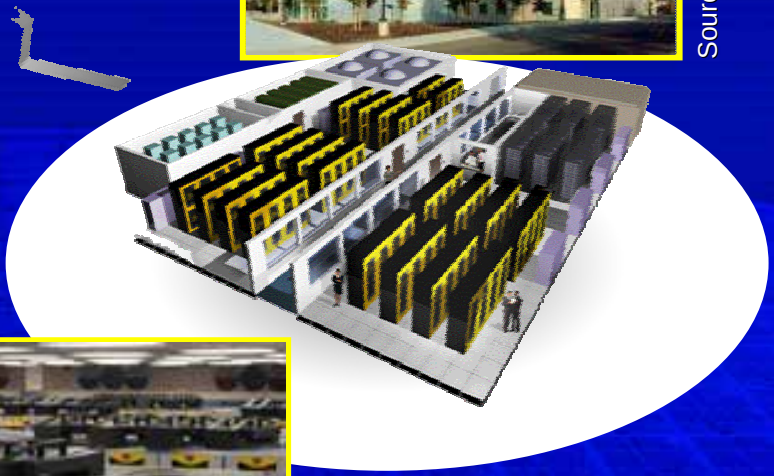
DATA CENTER "ENERGY LABEL"



ASC Purple

6 MWatt
100 TFlops goal
12K+ cpus – Power5
$230M

Source: LLNL



NASA Columbia

2 MWatt
60 TFlops goal
10,240 cpus – Itanium II
$50M

Source: NASA

intel

Power Efficiency

100W DELIVERED

20W

VR LOSS

45W

PSU LOSS

25W

UPS + PDU LOSS

190W SUPPLIED

TODAY'S TOTAL

52% EFFICIENT

100W DELIVERED

190W SUPPLIED

VR & PSU

TODAY	INTEL '10 GOAL
61% EFFICIENT	85% EFFICIENT
100W DELIVERED	100W DELIVERED
165W SUPPLIED	117W SUPPLIED

intel

Source: Intel Corporation

Nearly-Free FLOPS

NASA COLUMBIA EXAMPLE

TODAY

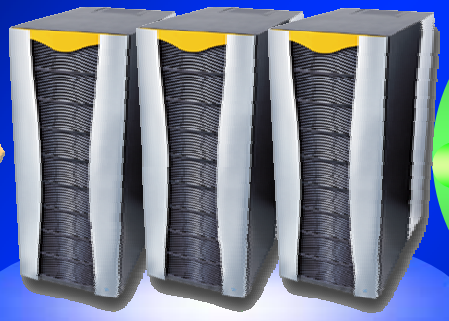
60% POWER EFFICIENCY

85 kW PDU
51kW DELIVERED
34kW LOST

Intel® x512

2010

85% POWER EFFICIENCY *



Intel® x512 + Intel® x256

90 kW PDU
77kW DELIVERED
13kW LOST

6% INCREASE IN POWER = 50% INCREASE IN PROCESSORS

intel

* POWER EFFICIENCY IMPROVEMENTS IN VRs & PSUs ONLY. UPSes & PDUs REMAIN CONSTANT.

Summary

- Great HPC advances made by industry
- Petaflops computing on the horizon
- Many cores, many threads coming
- Memory advances being made
- Optical development underway
- Power consumption being addressed across the power train spectrum

intel

Thank You!

谢谢